Masked Generative Distillation

Year: 2022 | Citations: 229 | Authors: Zhendong Yang, Zhe Li, Mingqi Shao, Dachuan Shi, Zehuan Yuan

Abstract

Knowledge distillation has been applied to various tasks successfully. The current distillation algorithm usually improves students' performance by imitating the output of the teacher. This paper shows that teachers can also improve students' representation power by guiding students' feature recovery. From this point of view, we propose Masked Generative Distillation (MGD), which is simple: we mask random pixels of the student's feature and force it to generate the teacher's full feature through a simple block. MGD is a truly general feature-based distillation method, which can be utilized on various tasks, including image classification, object detection, semantic segmentation and instance segmentation. We experiment on different models with extensive datasets and the results show that all the students achieve excellent improvements. Notably, we boost ResNet-18 from 69.90% to 71.69% ImageNet top-1 accuracy, RetinaNet with ResNet-50 backbone from 37.4 to 41.0 Boundingbox mAP, SOLO based on ResNet-50 from 33.1 to 36.2 Mask mAP and DeepLabV3 based on ResNet-18 from 73.20 to 76.02 mIoU. Our codes are available at https://github.com/yzd-v/MGD.